

Spatial variation of PM_{2.5} chemical species and source-apportioned mass concentrations in New York City

Kazuhiko Ito^{a,*}, Nan Xue^b, George Thurston^a

^a*Nelson Institute of Environmental Medicine, New York University School of Medicine, Tuxedo, NY, 10987, USA*

^b*Division of Biostatistics, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA*

Received 8 August 2003; received in revised form 13 February 2004; accepted 27 February 2004

Abstract

Particulate matter (PM) is a chemically non-specific pollutant, and may originate or be derived from different emission source types. Thus, its toxicity may well vary depending on its chemical composition. If the PM toxicity could be determined based on source types, the regulation of PM may be implemented more effectively. A large number of monitors began collecting PM less than 2.5 μm in diameter (PM_{2.5}) mass samples for subsequent chemical speciation starting 2000–2001 in the US. The data from this chemical speciation network can be useful for source-oriented evaluations of PM health effects. However, there are several issues that need to be considered in the analysis and interpretation of these data. One major issue is a monitor's representation of regional, sub-regional, and local air pollution exposures to the population in a city or metropolitan area. Because health outcomes in time-series air pollution epidemiological studies are aggregated over a wide geographical area, regional PM pollution may have smaller errors in exposure estimates than more spatially varying local pollution. However, the relative strength of association between source-apportioned PM and health outcomes may not be interpretable as the relative causal role of the source types. To our knowledge, there has not yet been a systematic and quantitative evaluation of this issue. In this study, we attempt to evaluate this issue by analyzing newly available PM_{2.5} speciation data from three monitors (a few miles apart) in New York City during 2001–2002. The strongest temporal correlations across the three monitors were found for the individual PM components that are related to secondary aerosols (e.g., S, NH₄). We also conducted source-apportionment of the data using absolute principal component analysis and positive matrix factorization. We identified four major source/pollution types: (1) secondary (largely regional) aerosols; (2) soil; (3) traffic-related; and (4) residual oil burning/incineration, in each of the three monitors. The estimated source-apportioned PM_{2.5} mass showed generally the highest monitor-to-monitor correlation for the secondary aerosol factor (r range: 0.72–0.93). The correlation for the more localized traffic-related factor was more variable (r range: 0.26–0.95). The estimated mean PM_{2.5} mass contributions by source/pollution type across the monitors varied least for the secondary aerosol factor. The extent of variability in the source-apportioned PM_{2.5} mass by the monitor was comparable to that from the difference due to the two source-apportionment techniques used. The implication of the results of our study is that a source-oriented evaluation of PM health effects needs to take into consideration the uncertainty associated with spatial representative of the species measured at a single monitor.

© 2004 Elsevier Ltd. All rights reserved.

*Corresponding author. Tel.: +1-845-731-3540; fax: +1-845-351-5472.

E-mail address: kaz@env.med.nyu.edu (K. Ito).

1. Introduction

Numerous studies have reported short-term associations between ambient air pollution concentrations and mortality and morbidity (e.g., see US EPA, 1996). Particulate matter (PM) was often implicated as the most significant predictor of the health outcomes among the air pollutants in these studies. However, PM is a chemically non-specific pollutant, and may originate or be derived from various emission source types. Thus, its toxicity may well vary depending on its source and chemical composition. If the PM toxicity could be determined based on source types, the regulation of PM may be implemented more effectively.

One natural progression of the PM health effects research is therefore to conduct a source-apportionment of PM using chemical speciation data (i.e., trace element, ions, organic and elemental carbons, etc.), and to examine the associations between source-apportioned PM (or underlying factor without mass attribution) and health outcomes, rather than with PM mass overall. There have been only a few studies that conducted such analysis (Özkaynak and Thurston, 1987; Özkaynak et al., 1996; Tsai et al., 2000; Laden et al., 2000; Mar et al., 2000). These studies have provided suggestive evidence that PM from certain combustion sources (i.e., secondary aerosols and traffic), but not other sources (e.g., soil), were associated with daily mortality. However, the availability of data sets that have collected appropriate speciation data has been limited to date, and the expected regional variability in source types makes it difficult to interpret these results in a straightforward manner.

Fortunately, a large number of monitors have now started collecting chemical speciation data from PM_{2.5} filters starting in 2000–2001 in the US PM_{2.5} chemical speciation network is one of the monitoring requirements set forth by the Federal Register (62 FR 38763), as part of the PM_{2.5} National Ambient Air Quality Standard (NAAQS) review completed in 1997 (US EPA, 1999). Each chemical speciation monitor measures: (1) PM_{2.5} particulate mass; (2) anions (sulfate, nitrate) and cations (particulate ammonium, sodium, and potassium) by ion chromatograph; (3) trace elements by energy dispersive X-ray fluorescence (EDXRF); and (4) total carbon including organic, elemental, and carbonate carbon by thermal optical analysis (US EPA, 1999). Each state is required to submit the speciation data to EPA's Aerometric Information Retrieval System (AIRS) three months after the collection of samples. Thus, the data from this chemical speciation network will provide a great research opportunity for source-oriented evaluations of PM health effects.

However, there are several outstanding issues that need to be taken into consideration when interpreting the results from source-apportionment and health effects

analysis of these PM_{2.5} speciation data. One major issue is the relative measurement error associated with each of the speciation components, and each of the source-types that are identified. The uncertainty associated with each measurement is estimated for certain chemical analyses (e.g., EDXRF). In fact, a source-apportionment technique, positive matrix factorization (PMF, Paatero and Tapper, 1993), uses such analytical uncertainty information to weight each observation. However, from an epidemiological point of view, there are other types of errors that may distort the potential associations between source-apportioned PM and the health outcome of interest. For example, in population-based time-series analysis, daily fluctuations of PM (or source-apportioned PM) are correlated with daily fluctuations of mortality or morbidity counts that are aggregated for the entire city. Often, the PM data come from a single monitor, or at most a few monitors per urban area. Most of the past source-apportionment analyses relied on data from a single monitor in a city. In such cases, it is possible that a single monitor does not represent each of the source types equally precisely and/or accurately for the population exposure. For example, regional secondary aerosols are usually more homogeneously distributed across a city than particles that originate from local mobile sources or point sources. In such cases, the relative strength of association between source-apportioned PM and health outcomes may not be interpretable as the relative causal role of the source types. To our knowledge, there has not yet been a systematic and quantitative evaluation of this issue. In this study, we attempt to evaluate this issue by analyzing data from three monitors in New York City that collected PM_{2.5} chemical speciation measurements during 2001–2002 study period.

2. Materials and methods

2.1. Data

The speciation data in New York City during 2001–2002 period were downloaded from New York State Department of Environmental Conservation (NYSDEC) Web site (<http://www.dec.state.ny.us/website/dar/baqs/pm25data.html>). There were three monitors that collected PM_{2.5} speciation data every-third-day: New York Botanical Gardens (NYBG) in Bronx, I.S. 52 in Bronx, and Queens College (QC) in Queens. Fig. 1 shows the locations of these three monitors in New York City. The distance between these three sites were 1.8 (NYBG–IS52), 4.7 (NYBG–QC), and 5.7 (IS52–QC) miles. The Botanical Gardens site used a Met One sampler, whereas the other two sites used R and P 2300 samplers. Data for the overlapping sampling period of April 2001–December 2002 were analyzed.

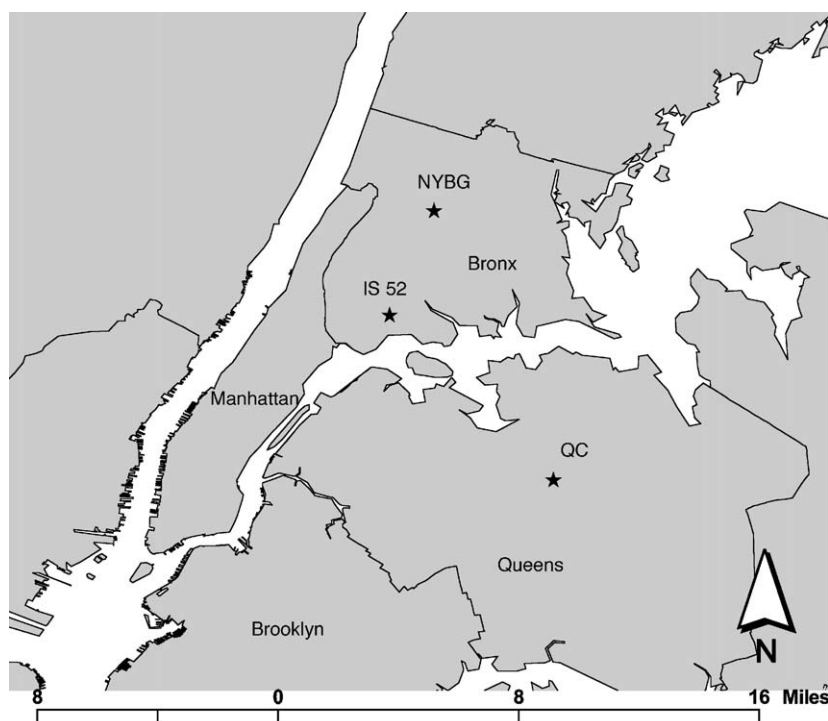


Fig. 1. Map of New York City and the three monitors.

The actual complete full observations (days with all species measured) during this period were 170 days at IS52, 171 days at QC, and 189 days at NYBG. The scheduled sampling days were the same for these monitors. Each chemical speciation monitor measured: (1) $PM_{2.5}$ particulate mass; (2) anions (sulfate, nitrate) and cations (particulate ammonium, sodium, and potassium) by ion chromatograph; (3) 48 trace elements by EDXRF; and (4) total carbon including organic, elemental, and carbonate carbon by thermal optical analysis. The analytical uncertainty values were not included in the original data from NYDEC or from the EPA's AIRS data that we also retrieved for the same monitors. However, the minimum detection limit (MDL) for each of the PM components was available from AIRS database, allowing us to compute the percentage above MDL for each series.

While the study period included 11, September, 2001 World Trade Center disaster, the every-third-day sampling of the speciation monitor missed the 11th and the highest impact day (12th September), and the data for the scheduled 13th and 16th September were missing from the three monitors' data sets. Based on a preliminary analysis of the daily speciation measurements a downtown Manhattan monitor near the World Trade Center and NYU speciation measurements (Thurston et al., 2003) and an examination of time-

series plots of speciation measurements at a monitor (Hunter College) in mid-town Manhattan (these data were not analyzed for this manuscript), the strong impact of World Trade Center collapse/fires was observed at Hunter College monitor on 12 September. While the impact of the demolition and continued burning were observed at the downtown monitor, its influence on measurements at the Hunter College monitor was minimal. Thus, we believe that the impact of the World Trade Center disaster is minimal on the three monitors' data that were analyzed in this study.

Another notable event during the study period was the 4 July (Independence Day in US) weekend, 7 July (Sunday), 2002, when high values of $PM_{2.5}$ and certain species (an apparent impact of a Quebec forest fire) were observed. We conducted analyses with and without these dates, but since the results were not qualitatively different, we present the results with these two days included.

2.2. Methods

The main objective of this analysis is to examine the difference in concentration of the speciated $PM_{2.5}$ constituents measured across three monitors, and to compare the source apportionment results from these three monitors. It should be noted that differences in

concentration of a given species between two monitors could be due to either: (1) actual difference in ambient concentration, or; (2) analytical measurement errors. The temporal correlation and difference in concentrations between a pair of monitors are influenced by both of these two components, but we did not attempt to separate out these components. However, we did examine how the monitor-to-monitor correlation was affected by the percentage above MDL, a crude indicator of the “signal-to-noise” ratio.

There are two components of any differences between measurements made at two monitors: one is the temporal correlation of the measurements; and the other is the quantitative difference in concentration. Thus, measurements made at two monitors can be perfectly correlated over time, but at the same time can have a large difference in concentration if one monitor always measures higher concentrations than the other monitor does. From an epidemiological point of view, the former is important for the strength of association, while the second is important for the effect size (e.g., risk per $\mu\text{g m}^{-3}$). Therefore, for each of chemical speciation data (56 species in total), temporal fluctuations in each pair of the three monitors were characterized in two ways: (1) temporal correlation; and, (2) coefficient of variation (C.V.) of the daily difference as a fraction of the overall mean (Evans et al., 1984).

Using each of these three monitors' data, source apportionment was first conducted using Absolute Principal Component Analysis (APCA) (Thurston and Spengler, 1985). Briefly, APCA employs a conventional (off the shelf) principal component analysis (PCA), with correlation matrix as input, followed by any factor rotation of choice. Then, for each of the principal components, the resulting scoring coefficients are used to compute a factor score for a hypothetical case where all the trace elements are zeros (i.e., a day without any pollution). This “offset” is then added to the all factor scores so that the factor scores are generally positive deviations from the zero pollution case. The idea is that, when the individual elements from the source are zeroes, so should be the composite mass contribution (i.e., factor score) from the source. For an illustrative example, see Thurston and Spengler (1985). Thus, in contrast to the other recently developed source apportionment techniques, PMF (Paatero and Tapper, 1993) or SAFER/UNMIX (Kim and Henry, 2000) in which positive mass concentration is used as a constraint for factor analytic solution, in APCA, an adjustment is made after the PCA and rotation so that mass contributions from sources are expected to, but not forced to be, positive (note that, depending on the scoring coefficients and the underlying elemental concentrations, the resulting APCA scores may take negative values, especially on very clean days).

We initially used APCA to analyze the data, but in response to reviewers' comments, we subsequently analyzed the data using PMF for comparison, as well. Note that, since we did not have the analytical uncertainty data for each observation, we used 5% of the measured value plus the minimum detection limit (to avoid near zero uncertainty at very low values) as estimates of the analytical uncertainty. We did not include all of the available PM trace constituents in the APCA analysis. The PM constituents that we could not use for identifying any specific source types were excluded. When redundant tracers were available, only one of them was included (e.g., SO_4 and S). Thus, the included species were: Al, As, Ba, Br, Cd, Ca, Cl, Cr, Co, Cu, Fe, Pb, Mg, Mn, Mo, Ni, P, K, Se, Si, Na, Sr, S, Ti, V, Zn, elemental carbon (EC), organic carbon (OC), NH_4 ion, and NO_3 ion. The criterion to decide the numbers of source factors to be retained depends on the objective of the study. While the common practice and program default in PCA appear to choose the “eigenvalue=1” cut-off (i.e., a variable's average “share” of contribution to total variance), it is possible that a factor with an eigenvalue less than 1 can significantly contribute to $\text{PM}_{2.5}$ mass. Thus, we initially explored the solutions in which factors with eigenvalues less than 1 were retained. However, those solutions did not provide any additional factors that could significantly contribute to $\text{PM}_{2.5}$ mass concentration. Therefore, with no other criteria to employ, the “eigenvalue=1” cut-off criterion was used to decide the number of factors to retain in this case. In order to explore a range of orthogonal to oblique rotations, we used the Harris–Kaiser rotation (Harris and Kaiser, 1964). We applied several sets of obliqueness, but with no changes in the interpretability of the resulting factor pattern, an essentially orthogonal rotation (Harris–Kaiser rotation with power=0.95) was chosen. $\text{PM}_{2.5}$ concentrations were then regressed on these factor scores. Resulting regression coefficients were used to compute daily contributions to $\text{PM}_{2.5}$ from each of the source types. In our PMF analysis, we used the same sets of PM species as those used in APCA analysis to make a comparison more direct. $\text{PM}_{2.5}$ concentrations were then regressed on the resulting factor scores from PMF. The statistical package, SAS (SAS Institute Inc., Cary, NC), was used to run PCA and regressions. We obtained PMF program from Dr. Paatero from University of Helsinki who wrote the program (Paatero, 1998).

To identify and “name” source/pollution type(s) for each factor is not straightforward. Except for fairly unambiguous source/pollution types such as soil (Al, Fe, Si, Ti) or secondary sulfate (NH_4 , SO_4), major portion of which in this region is transported sulfate from coal combustion, there is always a chance that a source is misidentified. Therefore, without an exhaustive confirmatory analysis, in this analysis we only suggest

source types for each factor. To aid in identifying source/pollution types, we relied on: (1) past source apportionment literature; (2) EPA's source profile repository software, SPECIATE (version 3.1); and (3) day-of-week pattern of the individual PM_{2.5} species and source-apportioned PM_{2.5}. Since we conducted APCA analysis first and went through the process of “naming” the factors, our PMF analysis was not an “independent” analysis. We identified similar factors and compared the estimated source-apportioned PM_{2.5} mass concentrations.

3. Results

Mean concentrations for PM_{2.5}, and key ions and trace elements for the three monitors are shown in Table 1. PM_{2.5} and species associated with secondary aerosols showed comparable levels across the three monitors. It can be seen that the secondary sulfate and

nitrate together (NH₄, SO₄, and NO₃) explain about 50% of the PM_{2.5} mass. The second largest constituent was OC (25–30%).

Fig. 2 shows temporal correlation for each pair of the three monitors for each of the 57 PM_{2.5} components, sorted by the smallest of the three correlations. The constituents of the secondary aerosols (i.e., S, NO₃, SO₄, NH₄) showed consistently high ($r > 0.9$) temporal correlations across three monitors. The main constituents of soil showed a range of correlation, from high (Al, $r > 0.8$) to moderate (Si, Fe, and Ti, with $r \sim 0.7$ –0.85). While OC, a major contributor to traffic related PM, showed high temporal correlations across three monitors, the other traffic related carbon fraction, EC, which is sometimes considered to be a good tracer for diesel, showed poorer correlations ($r = 0.30$ –0.43). We also computed the pair-wise C.V. (daily difference) for each pair of the three monitors, sorted by the largest of the three C.V.s (results not shown). Again, those with smallest concentration differences tend to be those

Table 1
Mean (standard deviation) of selected PM_{2.5} species from three monitors April 2001 to December 2002

	Units	NYBG ($n = 189$)		IS52 ($n = 170$)		QC ($n = 171$)	
PM _{2.5}	($\mu\text{g m}^{-3}$)	15.8	(9.8)	16.1	(10.1)	15.1	(10.2)
EC	($\mu\text{g m}^{-3}$)	1.3	(0.6)	1.1	(0.7)	0.7	(0.5)
OC	($\mu\text{g m}^{-3}$)	5.0	(3.5)	4.0	(3.1)	3.7	(3.3)
NH ₄	($\mu\text{g m}^{-3}$)	1.7	(1.4)	1.9	(1.6)	1.9	(1.5)
NO ₃	($\mu\text{g m}^{-3}$)	1.9	(1.8)	2.2	(1.9)	2.0	(1.7)
SO ₄	($\mu\text{g m}^{-3}$)	4.2	(3.3)	4.1	(3.1)	4.3	(3.5)
Al	(ng m^{-3})	20.4	(44.7)	14.4	(39.1)	13.4	(36.3)
As	(ng m^{-3})	1.1	(1.4)	0.9	(0.9)	0.9	(0.8)
Ba	(ng m^{-3})	33.1	(25.5)	18.6	(12.0)	19.3	(12.8)
Br	(ng m^{-3})	3.5	(2.7)	4.5	(3.4)	4.0	(2.7)
Cd	(ng m^{-3})	1.8	(3.3)	1.0	(1.5)	1.0	(1.4)
Ca	(ng m^{-3})	50.7	(30.8)	53.5	(25.3)	45.9	(27.5)
Cl	(ng m^{-3})	17.2	(67.1)	49.3	(183.0)	40.9	(217.2)
Cr	(ng m^{-3})	1.4	(1.4)	0.7	(0.6)	0.7	(1.0)
Co	(ng m^{-3})	0.8	(1.1)	0.7	(1.0)	0.5	(0.9)
Cu	(ng m^{-3})	4.4	(4.1)	4.0	(2.8)	3.6	(2.8)
Fe	(ng m^{-3})	106.9	(58.6)	117.6	(59.9)	105.1	(66.6)
Pb	(ng m^{-3})	6.0	(5.7)	6.9	(7.0)	5.1	(4.0)
Mg	(ng m^{-3})	8.0	(16.2)	6.7	(20.1)	5.9	(12.6)
Mn	(ng m^{-3})	2.5	(2.1)	2.1	(1.7)	2.3	(1.9)
Mo	(ng m^{-3})	1.4	(2.1)	0.8	(1.0)	0.8	(0.9)
Ni	(ng m^{-3})	24.2	(20.7)	18.4	(12.8)	14.3	(13.1)
P	(ng m^{-3})	6.3	(9.3)	3.7	(6.9)	3.0	(6.0)
K	(ng m^{-3})	48.6	(58.1)	52.1	(56.4)	49.6	(52.4)
Se	(ng m^{-3})	1.3	(1.5)	1.4	(1.6)	1.4	(1.5)
Si	(ng m^{-3})	97.9	(111.6)	107.1	(88.5)	102.5	(113.6)
Na	(ng m^{-3})	90.0	(110.2)	61.1	(77.6)	69.0	(85.8)
Sr	(ng m^{-3})	1.1	(1.6)	0.8	(1.0)	0.8	(0.9)
S	($\mu\text{g m}^{-3}$)	1.4	(1.0)	1.4	(1.0)	1.5	(1.0)
Ti	(ng m^{-3})	6.8	(4.9)	5.4	(4.1)	5.3	(3.9)
V	(ng m^{-3})	6.4	(3.6)	6.1	(4.0)	6.0	(4.0)
Zn	(ng m^{-3})	32.4	(23.9)	32.1	(21.3)	27.6	(21.4)

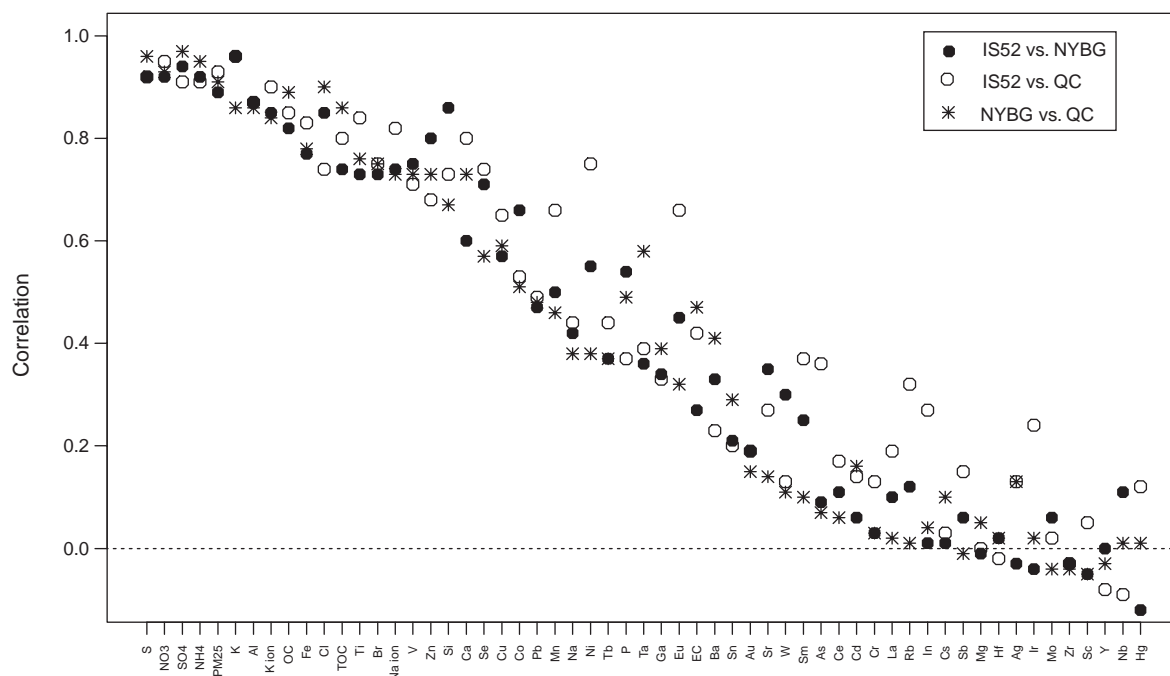


Fig. 2. Temporal correlation of PM components across 3 monitors, stored by the smallest of the three correlations.

associated with secondary aerosols (i.e., S, NO₃, SO₄, NH₄). Generally, the ranking of species for C.V.s was similar to that of temporal correlation, indicating that those species that correlated well were also closer in concentrations across sites.

We computed the median percentage of observations above the corresponding MDL for each species, and plotted them against the corresponding median monitor-to-monitor correlations in Fig. 3. It can be seen that many of the trace elements that showed low monitor-to-monitor correlations were also the ones with small percentages above the MDL. However, it is also clear that some PM species (e.g., EC, Ni, Mn, and Pb) showed low monitor-to-monitor correlations despite their high percentages above the MDL, suggesting that the observed discrepancies in concentrations and/or their temporal fluctuations are real.

Tables 2–4 show rotated factor patterns (i.e., factor loadings, or the correlation of each species with that factor) for the data from each of the three monitors using APCA. As mentioned in Methodology section, we also examined the cases with larger numbers of factors retained (with minimum eigenvalues smaller than 0.5), but those results did not provide additional factors that could significantly explain PM_{2.5}. Therefore, we chose the results with minimum eigenvalues larger than 1. This criterion resulted in 8 factors for all of the three monitors. It can be seen that, while the order of factors (i.e., eigenvalues after rotation) slightly varied, the three

monitors' results shared qualitatively similar factors. For example, a factor with high loadings on trace elements associated with crustal material (Al, Ca, Si, and Ti) was found for all three sites (Factor 2 for NYBG and QC; Factor 1 for IS52). The secondary sulfate factor, with high loadings on NH₄ and S, was also present for all three monitors. There was a factor with high loadings for OC, EC, and K for IS52 (Factor 5) and QC (Factor 4); for NYBG, OC and K (but not EC) shared a factor (Factor 4). The day-of-week pattern for EC at all three sites showed somewhat higher levels during the week than on the weekend, possibly reflecting diminished traffic (especially trucks), whereas OC's showed slightly higher levels on the weekend especially at NYBG site. The day-of-week pattern for this factor followed that of OC at NYBG and IS52, but the pattern was not clear at QC. Thus, while we considered this factor to be associated with traffic-related PM, there remained some uncertainty in naming this factor. Also, all the three monitors showed a factor that appeared to be a mixture of sources with signature elements for residual oil burning (Ni, V) and incineration (Cl, Pb, Zn) (Factor 1 for NYBG; Factor 3 for IS52; and, Factor 1 for QC). Nitrate loading was either with the oil burning/incineration factor (NYBG and QC) or split between oil burning/incineration factor and secondary sulfate factor (IS52). A factor with loadings for Na and Mg, present in all three monitors' data, may be sea salt (for example, see the sea salt composition in Seinfeld and

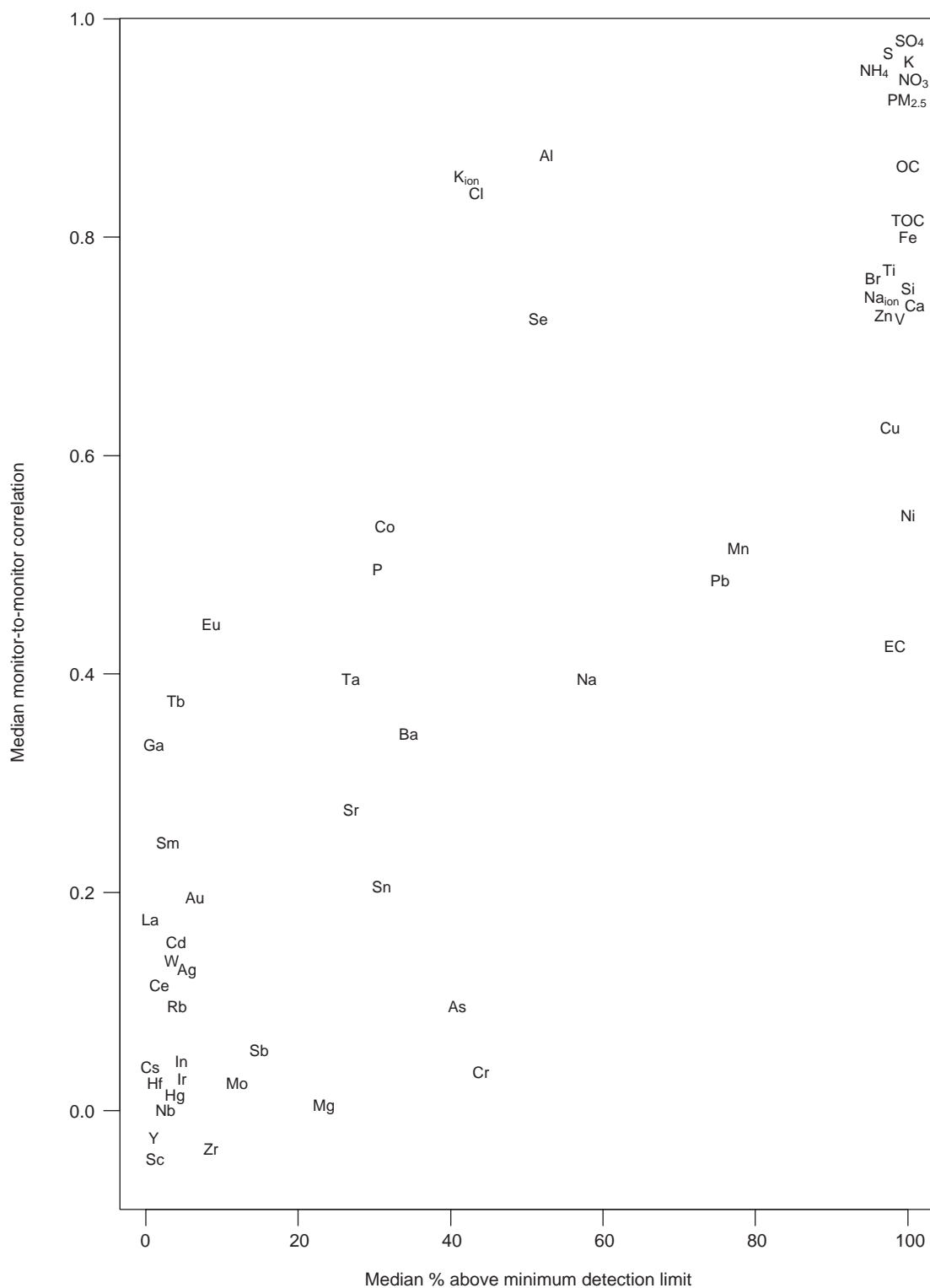


Fig. 3. Median (i.e., the middle of the three values) monitor-to-monitor correlation vs. median % above minimum detection limit of the PM_{2.5} species.

Table 2
Factor loadings for NYBG data using APCA

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
EC	0.74	0.17	0.11	0.08	0.12	−0.04	−0.07	0.11
OC	0.15	0.06	0.16	0.89	−0.05	−0.02	0.01	−0.04
NH ₄	0.40	0.17	0.78	0.13	−0.07	0.15	−0.01	0.05
NO ₃	0.74	0.03	0.32	0.08	0.00	0.24	0.03	0.12
Al	−0.14	0.84	−0.140	0.14	0.03	0.00	0.06	0.07
As	0.18	−0.05	0.00	−0.02	−0.03	0.02	−0.04	0.85
Ba	0.17	0.09	−0.04	0.07	−0.56	−0.41	0.13	0.21
Br	0.60	0.16	0.40	0.37	−0.02	0.13	−0.06	0.09
Cd	0.04	−0.05	−0.05	−0.11	0.15	0.55	0.28	−0.27
Ca	0.42	0.78	0.17	0.01	−0.02	0.09	−0.01	−0.02
Cl	0.55	−0.06	−0.02	0.09	−0.14	0.44	0.02	0.11
Cr	−0.11	−0.02	0.57	−0.13	0.08	−0.19	0.07	−0.07
Co	0.38	−0.10	−0.02	−0.01	0.59	0.23	0.08	0.25
Cu	0.29	0.18	0.26	0.04	−0.13	0.64	−0.18	0.14
Fe	0.38	0.77	0.28	0.09	0.01	−0.02	−0.07	−0.07
Pb	0.70	0.14	0.1	0.23	0.01	−0.06	−0.13	−0.31
Mg	−0.05	0.25	−0.13	0.08	0.10	−0.13	0.64	−0.21
Mn	0.04	0.46	0.29	0.36	0.36	−0.22	0.07	0.25
Mo	0.00	0.05	−0.07	−0.16	0.14	0.19	0.53	0.15
Ni	0.68	0.01	−0.16	−0.16	0.07	−0.01	0.27	0.05
P	0.13	−0.10	−0.08	0.01	0.87	−0.17	−0.08	−0.05
K	0.08	0.46	−0.02	0.80	0.03	0.06	0.05	0.02
Se	0.50	0.11	0.28	0.29	0.05	0.01	−0.06	0.35
Si	0.02	0.92	0.12	0.0	−0.02	0.0	0.02	−0.01
Na	−0.06	−0.09	0.09	0.10	−0.44	−0.08	0.71	0.02
Sr	−0.13	0.25	0.12	0.09	−0.04	0.62	−0.130	0.08
S	0.10	0.26	0.85	0.12	−0.06	0.07	−0.04	−0.02
Ti	0.04	0.84	0.12	0.16	−0.23	0.00	0.03	−0.06
V	0.77	0.17	0.25	0.03	0.02	−0.11	0.00	0.07
Zn	0.83	0.06	−0.02	0.09	0.20	0.05	−0.09	0.16

Pandis, 1998). A factor with loadings for Co and P were also present in all three monitors, but a close examination of the time-series of Co and P suggests that their correlation is in part due to a shared lack of signal in earlier part of the study period, which may be spurious. The remaining factors were unique to each monitor's results. In PMF analysis, we attempted runs with 6, 7, and 8 factor solutions for each of the three monitors' data. Factor loadings are not standard output of PMF, but we computed them by calculating the correlation between the original PM species and the 'G' matrix output from PMF (results not shown). Other than the choice of number of factors, we used default setting in PMF analysis. Since the 6, 7, and 8 factor solutions all showed the four major factors that are qualitatively similar to those observed in APCA results, we chose 8 factor solutions for further analysis. Each monitor's PM_{2.5} was then regressed on the eight factors simultaneously.

For all the three monitors, only four major factors (that are also common to the three monitors) contributed substantially (i.e., more than 1 μgm^{-3}) to the

PM_{2.5} mass in APCA. Table 5 shows a summary of the estimated mean contributions to PM_{2.5} from the four major source types for both APCA and PMF results. As expected, the least variability across sites was seen for secondary aerosols. There appears to be no clearly systematic difference between the two methods except the soil factor for which PMF consistently showed smaller estimated PM_{2.5} mass concentrations. Table 6 shows inter-correlation among the resulting source-apportioned PM_{2.5} for both APCA and PMF methods. Note that the APCA factors within each monitor were minimally correlated ($r < 0.1$) because of the choice of the nearly orthogonal rotations, whereas in PMF results, the correlation among factors over time at each site were more correlated (up to $r = 0.35$). In terms of correlation of the same source-type factors across monitors and methods, the estimated secondary sulfate factor was again most strongly correlated across monitors and methods (r range: 0.72–0.93). The soil factor (r range: 0.52–0.93) and residual oil/incineration factor (r range: 0.49–0.90) showed generally smaller correlation than secondary sulfate factor. The poorest

Table 3
Factor loadings for IS52 data using APCA

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
OC	0.10	0.16	0.06	−0.04	0.93	−0.08	0.02	0.04
NH ₄	0.20	0.87	0.21	−0.08	0.09	−0.06	−0.01	−0.01
NO ₃	0.02	0.56	0.59	0.11	0.12	0.06	0.02	0.16
Al	0.80	−0.09	−0.12	0.09	−0.01	0.42	0.01	0.07
As	−0.05	0.58	0.11	0.15	0.09	0.06	−0.22	0.20
Ba	0.20	0.04	0.15	−0.37	0.03	0.01	−0.60	0.12
Br	0.22	0.53	0.45	0.02	0.20	−0.02	−0.09	0.36
Cd	0.08	−0.10	0.08	−0.05	0.11	−0.15	0.65	0.04
Ca	0.69	0.28	0.0	0.09	0.35	−0.04	−0.02	0.06
Cl	0.00	0.10	0.79	−0.16	0.03	0.13	0.28	−0.10
Cr	0.05	0.38	0.01	0.26	−0.05	−0.01	0.43	0.11
Co	−0.03	0.03	0.41	0.75	0.04	0.00	0.06	0.14
Cu	0.33	0.60	0.29	−0.02	0.24	0.12	0.19	0.17
Fe	0.71	0.36	0.40	0.11	0.11	−0.10	−0.18	−0.02
Pb	0.08	0.26	0.33	0.14	0.21	−0.06	−0.07	0.63
Mg	0.10	−0.04	−0.09	0.00	−0.12	0.88	0.01	0.04
Mn	0.51	0.20	0.07	0.46	0.12	−0.18	−0.37	0.22
Mo	−0.07	0.18	0.12	0.43	0.07	0.42	0.31	−0.30
Ni	0.06	0.13	0.83	0.33	0.03	−0.04	−0.13	0.07
P	−0.05	−0.13	0.11	0.84	−0.06	−0.14	0.05	−0.01
K	0.51	0.01	0.00	0.01	0.71	0.00	0.10	0.24
Se	0.11	0.62	0.40	0.20	0.08	0.01	0.01	0.42
Si	0.85	0.23	−0.03	−0.07	0.10	0.00	0.14	0.02
Na	−0.05	−0.06	0.16	−0.25	−0.03	0.76	−0.12	−0.08
Sr	0.57	0.17	−0.01	0.26	0.19	−0.04	0.30	0.13
S	0.30	0.85	−0.05	−0.11	0.08	−0.11	−0.01	−0.08
Ti	0.88	0.16	0.09	−0.23	0.08	−0.04	−0.13	−0.04
V	0.24	0.45	0.57	0.16	0.07	0.01	−0.25	−0.04
Zn	0.15	0.13	0.69	0.31	0.15	−0.10	−0.07	0.28

correlation among the four major sources was found for the factor that we considered as traffic related (r range: 0.26–0.95), although some monitors showed high correlations between the methods. Fig. 4 shows time-series plots of the resulting source-apportioned PM_{2.5} mass concentrations for the four major source types for both APCA and PMF results for the QC monitor. It can be seen that the estimated PM_{2.5} mass concentrations for APCA and PMF generally concur well with occasional notable outliers (e.g., the traffic-related PM_{2.5} on July 2002). Note also that there are some negative PM_{2.5} values in the APCA results, although their magnitudes are small.

4. Discussion

The results from this study showed that the temporal correlation across monitors in NYC varied considerably (essentially zero to over 0.9) across individual PM_{2.5} species, indicating that the precision of population exposure estimates for specific elements can vary depending on the species. As expected, the constituents

of secondary aerosols showed consistently high ($r > 0.9$) temporal correlations across three monitors. Other PM_{2.5} species that are constituents of major source types (soil, traffic, oil burning, and incineration) showed low to high correlation ($r \sim 0.4$ –0.9) within the relatively short separation distances (~ 2 –6 miles) between monitors. For example, the temporal correlation for EC (sometimes considered to be a component of diesel exhaust) across three monitors ranged between 0.36 and 0.47. This indicates that daily fluctuations of the source type(s) represented by this species could not be precisely estimated for population exposure for a larger geographic boundary (i.e., the entire city). Thus, if a single monitor's or a few monitors' data are used to estimate the entire city's population exposure, then the potential health effects of individual PM species that have low monitor-to-monitor correlation such as EC would be masked or underestimated compared to PM species that have high monitor-to-monitor correlation (e.g., sulfate).

A composite pollution index (e.g., factor-analysis derived source-apportioned PM mass concentrations), rather than a individual PM species, may be more useful

Table 4
Factor loadings for QC data using APCA

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
EC	0.38	0.05	0.19	0.60	0.03	−0.09	0.27	0.16
OC	0.02	0.08	0.25	0.84	0.02	0.21	−0.11	−0.05
NH ₄	0.25	0.16	0.87	0.0	−0.03	−0.11	−0.04	−0.01
NO ₃	0.74	−0.01	0.44	0.04	0.09	0.01	−0.02	−0.16
Al	−0.12	0.89	−0.03	0.08	0.12	0.07	−0.08	−0.03
As	0.37	0.14	0.20	0.24	0.22	0.12	0.25	−0.33
Ba	0.09	0.11	0.01	−0.02	−0.24	0.16	0.75	0.03
Br	0.52	0.20	0.62	0.25	0.01	0.20	0.05	0.08
Cd	0.22	−0.06	0.0	0.29	−0.12	−0.11	−0.52	0.36
Ca	0.51	0.62	0.15	0.32	0.06	0.07	0.14	0.10
Cl	0.74	0.03	−0.03	0.08	−0.36	0.01	−0.31	0.14
Cr	0.07	0.00	0.04	−0.06	0.12	0.11	0.00	0.86
Co	0.71	0.01	−0.09	−0.01	0.0	−0.03	−0.18	−0.04
Cu	0.67	0.20	0.12	0.35	−0.18	−0.14	0.20	0.16
Fe	0.49	0.56	0.27	0.21	0.11	−0.09	0.40	0.13
Pb	0.60	0.16	0.1	0.13	0.19	0.06	0.16	0.11
Mg	−0.01	0.04	−0.15	0.36	0.03	0.82	0.00	0.04
Mn	0.20	0.40	0.39	0.00	0.47	0.13	0.39	0.06
Mo	0.29	−0.10	−0.04	0.28	0.39	−0.04	0.18	0.21
Ni	0.89	0.08	0.07	0.01	0.16	−0.01	0.04	−0.02
P	0.22	−0.04	−0.05	0.03	0.84	−0.10	−0.14	0.11
K	0.10	0.43	0.12	0.69	0.04	0.26	−0.08	−0.15
Se	0.60	0.10	0.58	0.12	0.07	0.04	0.04	−0.03
Si	0.11	0.78	0.19	−0.02	−0.13	−0.06	−0.05	−0.02
Na	0.06	0.00	0.00	−0.02	−0.09	0.88	0.13	0.06
Sr	0.26	0.37	−0.05	0.44	−0.03	30.00	0.23	0.00
S	−0.03	0.08	0.91	0.12	−0.04	−0.12	0.03	0.03
Ti	0.17	0.84	0.21	0.12	−0.09	−0.01	0.21	−0.03
V	0.62	0.13	0.53	0.04	0.06	0.13	0.24	−0.01
Zn	0.80	0.12	0.19	0.05	0.33	0.08	0.17	0.12

Table 5

Estimated mean PM_{2.5} contributions (standard error in parenthesis based on the PM_{2.5} regression standard error) and percentage from four common major factors identified in each of the three monitors using APCA and PMF (as matched with corresponding APCA factor for comparison)

	Factor A	Factor B	Factor C	Factor D
Suggested source(s)	Soil	Secondary sulfate	Traffic	Residual Oil/ incineration
Common species	Al, Ca, Fe, Si, Ti	NH ₄ , S	OC, EC (not at NYBG), K	Ni, V, NO ₃ / Cl, Pb, Zn
Factor # for each monitor's solution (see Tables 2–4)	NYBG: Factor 2 IS52: Factor 1 QC: Factor 2	NYBG: Factor 3 IS52: Factor 2 QC: Factor 3	NYBG: Factor 4 IS52: Factor 5 QC: Factor 4	NYBG: Factor 1 IS52: Factor 3 QC: Factor 1
<i>Est. Mean PM_{2.5} contribution (µgm^{−3})</i>				
NYBG	2.2 (0.3) 14%	5.6 (0.3) 35%	2.5 (0.1) 16%	4.2 (0.4) 27%
IS52	2.9 (0.4) 18%	6.7 (0.4) 42%	6.2 (0.4) 39%	1.8 (0.4) 11%
QC	1.6 (0.3) 11%	6.8 (0.3) 45%	5.5 (0.3) 36%	1.8 (0.2) 12%
<i>PMF results as matched to APCA factors</i>				
NYBG	0.4 (0.1) 3%	5.0 (0.2) 32%	3.2 (0.1) 20%	2.7 (0.2) 17%
IS52	0.8 (0.5) 5%	4.3 (0.4) 27%	3.5 (0.9) 22%	3.0 (0.5) 19%
QC	1.1 (0.5) 7%	6.0 (0.5) 40%	5.1 (0.8) 34%	2.2 (0.5) 15%

Table 6
Inter-correlation among source-apportioned PM_{2.5}: APCA and PMF results

	n.a.A	n.a.B	n.a.C	n.a.D	i.a.A	i.a.B	i.a.C	i.a.D	q.a.A	q.a.B	q.a.C	q.a.D	n.p.A	n.p.B	n.p.C	n.p.D	i.p.A	i.p.B	i.p.C	i.p.D	q.p.A	q.p.B	q.p.C	q.p.D
n. a. A	1.00																							
n. a. B	0.03	1.00																						
n. a. C	0.04	0.00	1.00																					
n. a. D	0.01	−0.01	−0.01	1.00																				
i.a.A	0.86	0.08	0.21	0.01	1.00																			
i. a. B	0.05	0.83	−0.02	0.21	0.02	1.00																		
i.a.C	−0.03	−0.02	0.80	0.06	0.02	−0.02	1.00																	
i.a.D	−0.09	−0.07	−0.07	0.79	−0.01	0.010	−0.02	1.00																
q. a. A	0.85	−0.03	0.27	−0.04	0.88	−0.05	0.11	−0.07	1.00															
q.a.B	0.12	0.78	0.14	0.21	0.13	0.83	0.10	0.03	0.07	1.00														
q.a.C	−0.03	−0.02	0.70	0.00	0.06	0.01	0.73	−0.10	0.03	0.00	1.00													
q. a. D	−0.01	0.02	−0.06	0.76	−0.01	0.21	0.03	0.76	−0.02	0.10	0.01	1.00												
n. p. A	0.91	−0.10	0.15	−0.11	0.82	−0.16	0.02	−0.13	0.89	−0.03	−0.03	−0.14	1.00											
n. p. B	0.28	0.82	0.11	−0.09	0.29	0.79	0.03	−0.25	0.21	0.81	0.08	−0.14	0.15	1.00										
n. p. C	−0.07	−0.07	0.95	0.02	0.09	−0.08	0.84	−0.07	0.14	0.09	0.75	−0.06	0.02	0.04	1.00									
n. p. D	−0.01	0.29	0.08	0.78	0.02	0.42	0.12	0.71	−0.02	0.43	0.03	0.75	−0.09	0.08	0.04	1.00								
i. p. A	0.61	0.07	0.08	0.21	0.67	0.00	−0.02	0.07	0.54	0.09	0.06	0.05	0.52	0.13	0.06	0.04	1.00							
i.p.B	0.18	0.78	0.09	−0.13	0.22	0.77	0.01	−0.26	0.13	0.72	0.04	−0.20	0.08	0.93	0.01	0.01	0.08	1.00						
i.p.C	0.15	0.21	0.26	0.32	0.10	0.28	0.65	0.25	0.10	0.28	0.34	0.35	0.04	0.17	0.27	0.34	0.09	0.17	1.00					
i.p.D	0.02	0.39	0.06	0.70	0.03	0.52	0.11	0.49	−0.02	0.53	0.00	0.56	−0.09	0.21	0.05	0.85	0.17	0.13	0.35	1.00				
q.p.A	0.83	−0.07	0.25	−0.14	0.84	−0.13	0.11	−0.17	0.93	−0.03	0.05	−0.13	0.93	0.15	0.12	−0.07	0.49	0.07	0.07	−0.07	1.00			
q.p.B	0.19	0.82	0.10	0.00	0.20	0.82	0.05	−0.18	0.11	0.91	0.05	−0.09	0.04	0.93	0.03	0.21	0.08	0.87	0.21	0.34	0.03	1.00		
q.p.C	0.13	0.12	0.38	0.25	0.15	0.26	0.42	0.12	0.05	0.25	0.72	0.27	0.00	0.18	0.39	0.27	0.16	0.10	0.47	0.24	0.03	0.20	1.00	
q.p.D	0.01	0.09	0.06	0.76	−0.01	0.28	0.13	0.70	0.01	0.27	0.03	0.85	−0.07	−0.09	0.05	0.90	0.00	−0.17	0.35	0.77	−0.05	0.03	0.27	1.00

The column/row notation: (1).(2).(3): (1)is monitor: 'n' for NYBG; 'i' for ISS2; and 'q' for QC; (2) is method: 'a' for APCA; 'p' for PMF; (3) is the source type A, B, C, and D, as noted in Table 5. The numbers in bold indicate correlation among the same factor category.

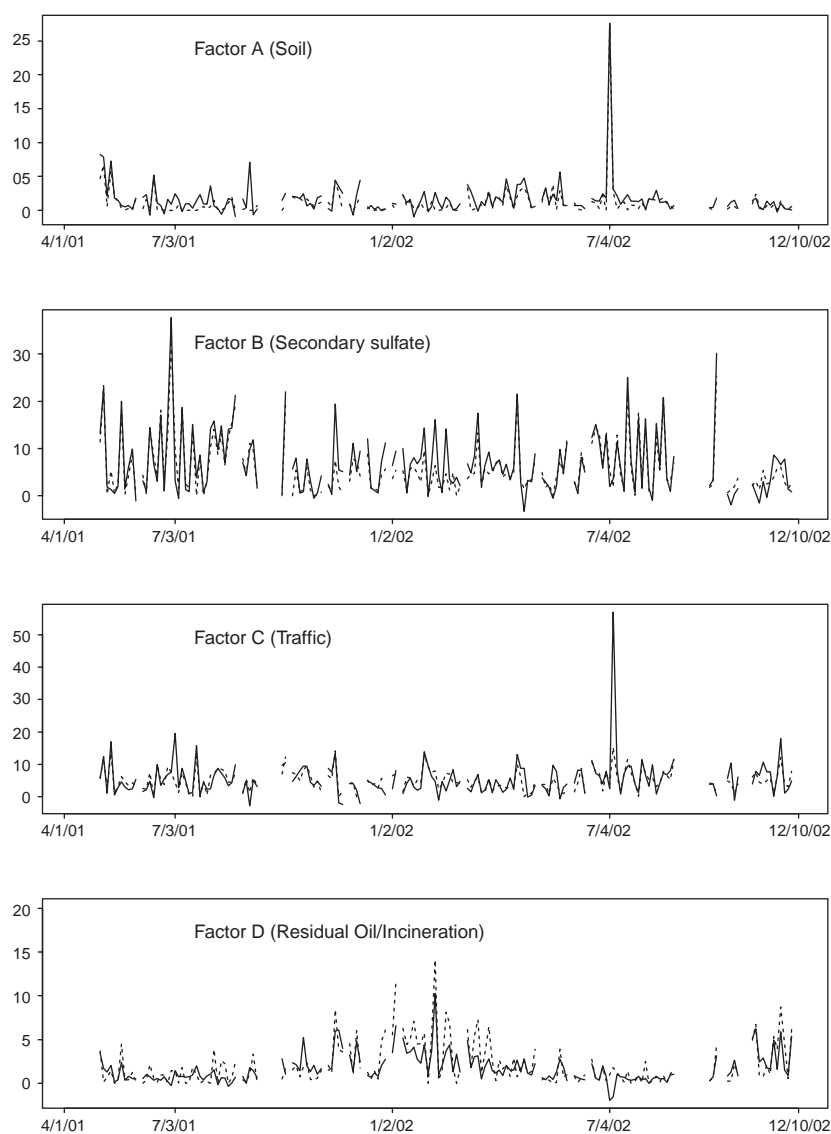


Fig. 4. Time-series plots of source-apportioned $\text{PM}_{2.5}$ ($\mu\text{g m}^{-3}$) by APCA (solid line) and PMF (dotted line): QC data.

in conducting source-oriented evaluation of PM health effects, but several complications of this approach need to be considered. The monitor-to-monitor correlation of the source-apportioned $\text{PM}_{2.5}$ mass concentrations for the four major source types were generally moderately high ($r > 0.7$), but varied from the consistently high correlation for the secondary aerosol factor (r : range 0.72–0.93) to the highly variable correlation for the “traffic-related” factor (r : range 0.26–0.95). The variable range of monitor-to-monitor correlation for the “traffic-related” factor also is an indication of larger uncertainty in naming this factor “traffic-related” for the exposure estimate for the entire city. The variable range is also expected from the lower correlation of this factor’s

constituent, EC. For the secondary aerosols, the range of correlation (0.72–0.93) meant that monitor-to-monitor correlation for the source-apportioned $\text{PM}_{2.5}$ was lower than the correlation for the individual tracer species for this source (NH_4 and S). Because a factor (or a principal component) is a linear combination of individual z -transformed species, the resulting factor may be noisier than the individual tracer species. In such cases, it may be more desirable to rotate the solution by targeting towards a single known tracer species. On the other hand, for other source types, it is also possible that a factor analysis can “extract” out an underlying source “signal” factor several tracer species that are individually noisy. This issue also complicates the use of

source-apportioned PM in the health effects analysis. If one species in the chemical mixture of PM were the causal agent, then source-apportioned PM (which is “noisy” due to the influence from other species) would not predict the adverse health effects as precisely as the responsible specie would. On the other hand, if a certain mixture of PM from a particular source were more toxic than its individual constituents alone, then the source-apportioned PM would be better correlated with health outcomes. This issue needs further investigation.

While the monitor-to-monitor correlation of source apportioned PM_{2.5} were often comparable across the four major source types found in this study, the mean mass contribution estimated for each source types varied by up to several fold (worst for soil, for which estimated mass concentrations were relatively low), except in the case of the secondary aerosols. The implication is that, except for secondary aerosols, if these source apportioned PM were used in time-series analysis of mortality or morbidity data, the absolute health risk estimates (per unit mass concentration) for each source type could vary by several fold, depending on which monitor's data were used. This is an issue that needs to be closely examined, especially when the absolute risk estimates per unit mass concentration for each source types are compared across different cities in which only a monitor or a few monitors are used. Such estimates need to take into consideration the variability of source-apportioned PM levels within that city.

We initially used APCA to analyze the data, and later also analyzed the same data with PMF. We could have also used UNMIX (e.g., see Lewis et al., 2003), but our schedule did not permit us to do so. Ramadan et al. (2000) summarized disadvantages of PCA-based analysis including its inability to incorporate uncertainty information and to address below-detection-limit data. Poissant (1994) also pointed out, through an example, that APCA might bias results. In the analysis of real data, if we attempt to estimate the population exposure to PM source types, there are several factors that can lead to differing solutions including: (1) the difference in model; (2) the options chosen within a model; and, (3) the choice of a monitor (where multiple monitors are available). We could not conduct a comprehensive analysis to determine relative importance of these factors, but we could describe the magnitude of variability in source-apportioned PM concentration estimates due to the difference in the model and the choice of monitors. The results generally suggest that variability of results in the mean mass contributions between the APCA and PMF approaches were roughly comparable to the variability due to the location of monitors. Since our PMF analysis did not take a full advantage of the use of actual analytical uncertainty of measurements, this was not a full comparison of the two approaches. Nevertheless, these two approaches did not

show a substantial or systematic difference in results. Overall, similar results and conclusions were achieved with either method.

The implication of the results of our study is that a source-oriented evaluation of PM health effects needs to take into consideration the uncertainty associated with spatial representative of the species measured at a monitor. It is possible that associations between a source-type and health outcome is distorted or not detected due to the error associated with the estimation of exposure for that source-type. The extent of such errors can be estimated when multiple monitors exist in a city. We suggest that a comprehensive assessment be conducted to investigate the relative uncertainty associated with each source type using the newly available PM_{2.5} chemical speciation network.

Acknowledgements

This research was supported by US EPA STAR grant (R827997010); US EPA Particulate Matter Health Research Center (R827351); and NIEHS Environmental Health Center (ES00260).

References

- Evans, J.S., Cooper, D.W., Kinney, P.L., 1984. On the propagation of error in air pollution measurements. *Environmental Monitoring and Assessment* 4, 139–153.
- Harris, C.W., Kaiser, H.F., 1964. Oblique factor analytic solutions by orthogonal transformations. *Psychometrika* 29, 347–362.
- Kim, B.-M., Henry, R.C., 2000. Application of the SAFER Model to Los Angeles PM₁₀ Data. *Atmospheric Environment* 34, 1747–1759.
- Laden, F., Neas, L.M., Dockery, D.W., Schwartz, J., 2000. Association of fine particulate matter from different sources with daily mortality in six U. S. cities. *Environmental Health Perspectives* 108, 941–947.
- Lewis, C.W., Norris, G., Henry, R., 2003. Source apportionment of phoenix PM_{2.5} aerosol with the unmix receptor model. *Journal of the Air and Waste Management Association* 53, 325–338.
- Mar, T.F., Norris, G.A., Koenig, J.Q., Larson, T.V., 2000. Associations between air pollution and mortality in Phoenix 1995–1997. *Environmental Health Perspectives* 108, 347–353.
- Ozkaynak, H., Thurston, G.D., 1987. Associations between 1980 U.S. mortality rates and alternative measures of airborne particle concentration. *Risk Analysis* 7, 449–460.
- Ozkaynak, H., Xue, J., Zhou, H., Raizenne, M., 1996. Associations Between Daily Mortality and Motor Vehicle Pollution in Toronto Canada Department of Environmental Health. Harvard University School of Public Health, Boston, MA.

- Paatero, P., Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. *Chemometrics and Intelligent Laboratory Systems* 18, 183–194.
- Paatero, P., 1998. User's guide for Positive Matrix Factorization programs PMF2 and PMF3. Available from <ftp://rock.helsinki.fi/pub/misc/pmf/>.
- Poissant, L., 1994. A practical demonstration of the absolute PCAS bias. *Atmospheric Environment* 28, 2129–2134.
- Ramadan, Z., Song, X.H., Hopke, P.K., 2000. Identification of Sources of Phoenix Aerosol by Positive Matrix Factorization. *Journal of the Air and Waste Management Association* 50, 1308–1320.
- Seinfeld, J.H., Pandis, S.N., 1998. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. Wiley, New York.
- Thurston, G.D., Spengler, J.D., 1985. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. *Atmospheric Environment* 19, 9–25.
- Thurston, G.D., Maciejczyk, P., Lall, R., Hwang, J.S., Chen, L.C., 2003. A source apportionment of fine particulate matter in lower Manhattan following the WTC disaster. Abstract # P13-2, presented at the 2003 AAAR PM Meeting: Particulate Matter: Atmospheric Sciences, Exposure and the Fourth Colloquium on PM and Human Health, March 30–April 2, 2003, Pittsburgh, PA.
- Tsai, F.C., Apte, M.G., Daisey, J.M., 2000. An exploratory analysis of the relationship between mortality and the chemical composition of airborne particulate matter. *Inhalation Toxicology* 12 (Suppl.), 121–135.
- US EPA. 1996. Air quality criteria for particulate matter. EPA/600/P-95/001.
- US EPA. 1999. Particulate matter (PM_{2.5}) speciation guidance final draft, 1st Edition. US EPA Monitoring and Quality Assurance Group. Emission, Monitoring, and analysis Division, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711.